ChatGPT in Healthcare: A Taxonomy and Systematic Review

Abstract

The recent release of ChatGPT, a chat bot research project/product of natural language processing (NLP) by OpenAI, stirs up a sensation among both the general public and medical professionals, amassing a phenomenally large user base in a short time. This is a typical example of the 'productization' of cutting-edge technologies, which allows the general public without a technical background to gain firsthand experience in artificial intelligence (AI), similar to the AI hype created by AlphaGo (DeepMind Technologies, UK) and self-driving cars (Google, Tesla, etc.). However, it is crucial, especially for healthcare researchers, to remain prudent amidst the hype. This work provides a systematic review of existing publications on the use of ChatGPT in healthcare, elucidating the 'status quo' of ChatGPT in medical applications, for general readers, healthcare professionals as well as NLP scientists. The large biomedical literature database PubMed is used to retrieve published works on this topic using the keyword 'ChatGPT'. An inclusion criterion and a taxonomy are further proposed to filter the search results and categorize the selected publications, respectively. It is found through the review that the current release of ChatGPT has achieved only moderate or 'passing' performance in a variety of tests, and is unreliable for actual clinical deployment, since it is not intended for clinical applications by design. We conclude that specialized NLP models trained on (bio)medical datasets still represent the right direction to pursue for critical clinical applications.